Nucleic Acids Research, Vol. 19, No. 14 3979–3986

Effect of intron A from human cytomegalovirus (Towne) immediate-early gene on heterologous expression in mammalian cells

Barbara S.Chapman*, Richard M.Thayer[1], Karen A.Vincent[2,+] and Nancy L.Haigwood[2]
Biochemistry, [1]Product Management and Regulatory Affairs and [2]Virology, Chiron Corporation, 4560 Horton Street, Emeryville, CA 94608, USA

Received February 15, 1991; Revised and Accepted June 10, 1991 GenBank accession no. M60321

ABSTRACT

A 2.4 kb fragment of hCMV (Towne strain), containing the 5′ end of the major immediate-early gene, has been cloned, sequenced, and used to construct a series of mammalian cell expression plasmids. The effects of regulatory regions present on this fragment were assessed using human glycoproteins as reporter molecules. We compared secreted levels of Factor VIII, t-PA, and HIV-1 envelope glycoproteins in cells transfected with plasmids in which intron A of the immediate-early gene was present or absent. Secretion of several glycoproteins was significantly higher when cells were transfected with intron A-containing plasmids. Mutation of three basepairs in the strong nuclear factor 1 (NF1) binding site in intron A led to reduced transient expression levels, but not to the level observed in the absence of intron A. Reduced expression from NF1 mutant plasmids was roughly correlated with reduced binding *in vitro* of NF1 proteins to a synthetic oligonucleotide containing the mutation. The evidence indicates that sequences in intron A positively regulate expression from the hCMV immediate-early enhancer/promoter in transformed monkey kidney cells.

INTRODUCTION

The immediate-early Mr 72,000 protein gene of human cytomegalovirus (hCMV) is transcribed from one of the strongest enhancer/promoters known (1,2,3). The enhancer appears to be active in a broad range of host cell types (1,4,5), and in comparison with various viral promoters, it has been shown to be the strongest in transient expression systems (4,6,7). The hCMV IE1 enhancer/promoter has proven a versatile source of transcriptional signals for expression of heterologous proteins (5,7,8,9,10,11).

Expression of the hCMV major immediate-early (IE1) gene is controlled by a complex of regulatory regions, which includes upstream sequences that specifically modulate expression in cell types (12,13), a cluster of binding sites for nuclear factor 1 (NF1) (3,14,15), and a complex enhancer (1,3). The transcribed region of the gene contains four exons and three introns (2,16). The largest of the introns, intron A, occurs within the 5′ untranslated region of the gene, where the strongest of five NF1 binding sites is located (14). While the enhancer region of this gene has been carefully studied, the possibility of regulatory signals in transcribed sequences has not been addressed until recently (17).

To examine hCMV IE1 regulatory signals in the transcribed region, we chose to express heterologous glycoproteins, rather than synthetic reporter molecules, in a series of expression plasmids containing hCMV IE1 DNA. In defining sequences of the hCMV IE1 region that produce the highest levels of glycoprotein secretion in a transient expression system, we show that intron A contains elements that enhance expression in transformed monkey kidney cells.

MATERIALS AND METHODS

Construction of recombinant plasmids

All plasmids containing hCMV (Towne) sequences were derived from pRL103a, a 21 Kb *Hin*d III (*Hin*d III C) fragment of hCMV (Towne) cloned in pBR322 (18). *E. coli* host strains having strong RecA phenotypes were used for all cloning steps (19). A 5.5 Kb *Hin*d III-*Bam*HI fragment of pRL103a, containing the 5′ end and upstream sequences of the major immediate-early gene, was subcloned into pBR322 for restriction mapping and sequencing. For cloning steps requiring site-directed mutagenesis, restriction endonuclease fragments were subcloned into M13 bacteriophage vectors (20), modified by synthesis of new (+) strands using oligonucleotide primers (21), and completely sequenced before replacement into the expression plasmids (22). The complete sequence of relevant DNA from pRL103a is shown in Fig.1.

A backbone plasmid (pSVori) was constructed to include the SV40 origin of replication and early region polyadenylation signal, *E. coli* origin of replication and ampicillin resistance gene, and restriction sites for inserting hCMV IE1 fragments (Fig.2).

* To whom correspondence should be addressed

+ Present address: Howard Hughes Memorial Institute, Beckman Center 253, Stanford University School of Medicine, Stanford, CA 94305, USA

pSVori is composed of a 1.5 Kb *Pvu* I to *Pst* I fragment of pSVT-2 (23) and a 0.7 Kb *Sal* I to *Pvu* I fragment of pSV7d (24).

5′ flanking and 5′ untranslated sequences of the hCMV Towne IE1 region are shown schematically in Fig.3A. A 1636 bp fragment of the CMV IE1 5′ sequence starting at the 3′ *Ssp* I site, including intron A and retaining two of the five NF1 binding sites, was created as follows by site-directed mutagenesis (Fig.3B). A *Sal* I site was installed 9 bp before the initiation codon. The resulting DNA was cut with *Ssp* I and *Sal* I, then ligated into the *Eco*RI (filled) and *Sal* I sites of pSVori to produce an expression vector called pCMV6 (Fig.2). A fragment beginning at the *Ssp* I site and lacking intron A was prepared by site-directed mutagenesis using a 63 base synthetic oligonucleotide to precisely delete intron A and create a *Sal* I site 9 bp before the initiation codon (Fig.3C). This deletion fragment was then cloned into pSVori to make pCMV5 (Fig.2). Finally, the NF1 binding site in intron A was altered by site-directed mutagenesis to the sequence TGGCAACTTGCCAA. This change of three bp in the spacer region produces an NF1 consensus sequence found in a middle repetitive element of the mouse genome, which has been reported not to bind NF1 (14). After mutagenesis, the entire clone was re-sequenced to verify that no other changes had been made. The fragment (Fig.3D) was used to make pSVori derivative pCMV6nf1.

Expression plasmids for human t-PA were constructed using the *Sal* I fragment from pSV7tPA3 (25) encoding the complete human t-PA plus 99 bp of 5′ untranslated DNA. This fragment was inserted into CMV IE1 plasmids pCMV5 (without intron A) and pCMV6 (with intron A). Expression plasmids for HIV-1 were of four types. The first included the entire gp160 envelope gene with the homologous signal sequence from HIV-SF2, on a *Sac* I (nt 5555) to *Xho* I (nt 8460) fragment. In the second type, the envelope gene was mutagenized at the processing site for gp120-gp41 cleavage, adding a stop codon following Arg509, to produce a secreted gp120 with the homologous HIV signal. In the third and fourth types, the 5′ end of the envelope gene was modified to place an *Nhe* I site upstream of Glu31 for replacement of the homologous signal of gp120 and gp160 with the heterologous human t-PA signal, modified to bear an *Nhe* I site at amino acid Ser23 to make a 25 amino acid signal sequence ending in AlaSer.

Sequence analysis

DNA fragments to be sequenced were subcloned into M13 bacteriophage vectors (20). Sequencing was carried out using the dideoxynucleotide chain termination method (22), and compressions resolved through the use of 7-deaza-2′-deoxy-GTP (26). The complete sequence of both DNA strands was determined by preparing single-stranded M13 templates and sequencing with a series of oligonucleotide primers synthesized on an Applied Biosystems 380A synthesizer. Data were analyzed using Analysis Release Package software (27).

Cell culture

Transient expression experiments were carried out using COS7 cells (monkey kidney cells transformed with SV40 T-antigen)(28) maintained in Dulbecco's modified Eagle's medium (DME) supplemented with 10% fetal bovine serum. COS7 cells were transfected with CsCl-banded plasmid DNA using a modification of a DEAE-Dextran/chloroquine diphosphate procedure (29,30). Monkey kidney cells carrying a temperature-sensitive SV40 T-antigen gene (ts COS) were maintained and transfected as were the COS7 cells (31). Temperature sensitive (ts) COS cells were co-transfected with pSV2-neo (32) and selected in 400 μg/mL G418. Negative control cell lines were transfected with pSV2-neo alone.

Assays for secreted glycoproteins

Concentrations of Factor VIII (FVIII) glycoproteins secreted by transfected cells were determined using the COATEST activity assay (KabiVitrum), a sandwich ELISA for FVIII gp80 light chain (33), and a solid phase radioimmunoassay for FVIII gp92 heavy chain (34).

Recombinant human t-PA in the culture medium was detected using an antigen capture ELISA (Imubind™, American Diagnostica) calibrated with human t-PA (NIBSC no. 83151). Recombinant HIV-1 glycoproteins were detected using an antigen capture ELISA developed at Chiron specific for viral or recombinant HIV-SF2 envelope. The capture antibody was high titer polyclonal goat serum raised against yeast-derived recombinant gp120 (available from the NIAID AIDS Research and Reference Reagent Program, January 1991 catalog no. 385). The detecting antibody was high titer rabbit polyclonal serum raised against the same recombinant antigen (1:100). Detection was via horse-radish peroxidase conjugated goat anti-rabbit (1:3000, Tago code 6400). The assay was calibrated with purified recombinant gp120 from HIV-SF2 produced in mammalian cells (35) (available from the NIAID AIDS Research and Reference Reagent Program, January 1991 catalog no. 386).

Total cytoplasmic RNA was extracted from permanent temperature sensitive (ts) COS cell lines (36). Equivalent amounts were electrophoresed through formaldehyde agarose gels and transferred to nitrocellulose. The *Sal* I insert fragment from pCMV6a120SF2 was labeled with ^{32}P by nick translation, then hybridized to immobilized RNA under standard conditions (36).

In vitro NF1 binding assay

Nuclear extracts containing NF1 proteins were prepared from COS7 cells as described (15). Oligonucleotides were made on an Applied Biosystems 380B DNA synthesizer. The 26 base NF1-V strand (TCTATTGGCTATATGCCAATACTCTG) and NF1-mutV strand (TCTATTGGCAACTTGCCAATACTCTG) were labeled by hybridizing a primer (CAGAGTATTGGC) and extending the duplex using the large fragment of DNA polymerase I (37). Ten fmol of each ^{32}P-labeled oligonucleotide were incubated separately with crude nuclear extract in 20μL mixtures and analyzed on non-denaturing polyacrylamide gels as described (15).

RESULTS

Sequence analysis of the 5′ region of hCMV (Towne) IE1 gene

The 5′ region of the hCMV (Towne) major immediate-early gene was subcloned into pBR322 as a 5.5 Kb *Hin*d III to *Bam*HI restriction fragment from pRL103a (18). The sequence shown in Fig. 1 was obtained using the M13 universal primer and internal oligonucleotide primers with three overlapping M13 subclones of the following restriction endonuclease fragments: *Pst* I to *Sac* I (nt 1−1125), *Ssp* I to *Pvu* II (nt 458−2356), and *Hinc* II (nt 1075−1975).

The sequence shown in Fig. 1 extends the known sequence of the hCMV (Towne) IE1 gene by 533 bases on the 5′ side and provides the complete sequences of introns A and B. When

aligned with the more extensively studied hCMV strain AD169 (1,14,38), the genes are 97% homologous in non-coding and coding DNA.

The 5′ end of the hCMV (Towne) IE1 gene can be divided into five regulatory regions. The first approximately 400 bp contains a tissue-specific modulator, thought to extend from nt 1 to ~365 (12). This region contains a potential stem-loop structure in the form of two 30 bp inverted repeats (nt 119–148 and 183–211) separated by 31 bp (Fig.1). Next there is a cluster of four binding sites for nuclear factor 1 (NF1) between nt 372 and 530 (14,15). The first site matches the consensus sequence TGGC/A(N)$_5$GCCAA required for optimal NF1 binding, but the other three vary at one or more positions. The function of these clustered NF1 sites has not been established, but NF1 binding in concert with other transcription factors may activate transcription from a downstream promoter (39).

The most important regulatory region is a complex enhancer, which spans nt ~600 to ~1081 (1,2,3). Many potential binding sites for well-characterized transcription factors, including Sp1 (CCGCCC), CRE/ATF (TGACGTCA) and NF-κB (GGGRC/A/TTYYCC) are present. Since these transcription factors are common to diverse cell types, the presence of clustered sequence motifs probably accounts for the broad spectrum and high level of activity observed for the hCMV IE1 enhancer. A strong polymerase II promoter extends from nt 1082 to the transcriptional start site at 1144 (16). Nucleotides 1082 to 1086

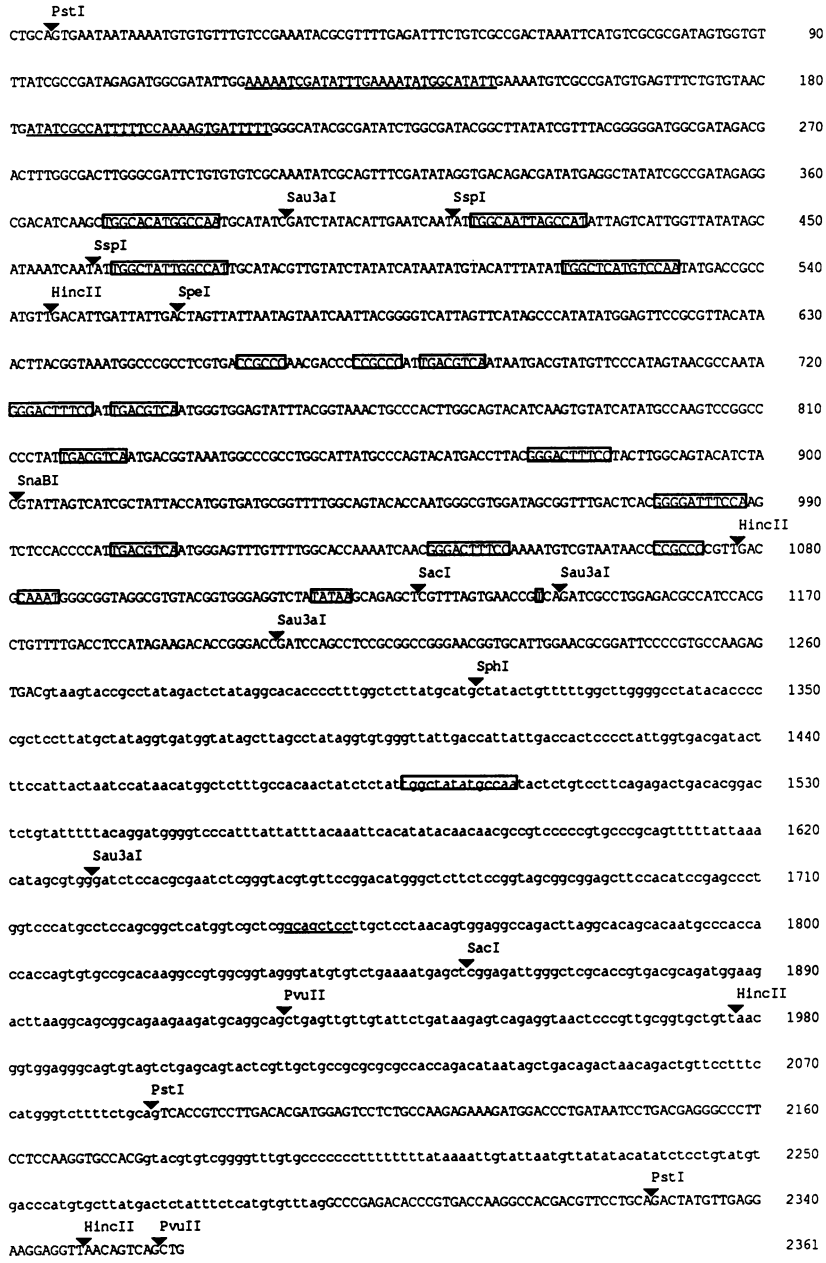

Fig. 1. Nucleotide sequence of the 5′ region of the major immediate-early gene of human cytomegalovirus (Towne). Bases are numbered starting with the *Pst* I site 1144 nucleotides 5′ of the transcriptional start site for the IE1 gene. Inverted repeat sequences are underlined. Binding sites for transcription factors are boxed. Introns A and B are shown in lowercase letters. GenBank accession number is M60321.

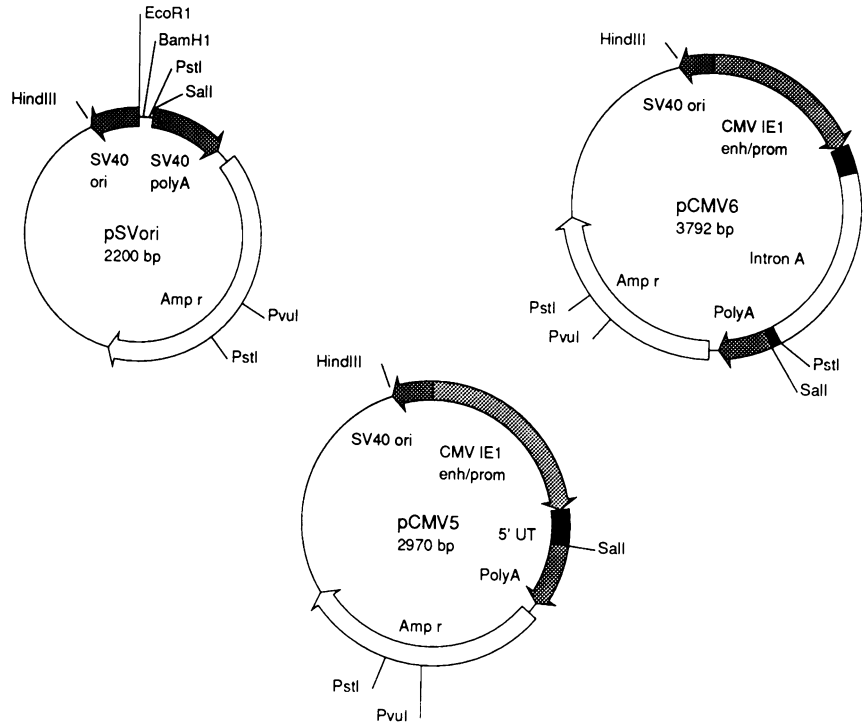

Fig. 2. Maps of plasmids pSVori, pCMV5 and pCMV6. The construction of pSVori is described in the text. Expression vectors were prepared by inserting fragments containing the CMV IE1 enhancer/promoter with intron A (pCMV6) or without intron A (pCMV5).

form a non-consensus CTF binding site (CCAAT-box motif) at position −62 relative to the start of transcription. At −27 (nt 1117 to 1122) there is a TATAA box ($TF_{II}D$ binding site). Transcription is initiated within a CREB (cyclic AMP responsive element binding protein) sequence (Fig.1).

A fifth regulatory region of the hCMV IE1 gene appears to be within transcribed sequences. Exon A (121 bp) is followed by intron A (823 bp) and exon B, in which translation is initiated (Fig.1). Exon A contains an 18 bp element homologous to that shown to regulate the level of transcription of the CMV immediate-early promoter (17). Searching intron A reveals no consensus motifs for Sp1, CRE/ATF, NF-κB, or CREB. However, a perfect consensus site for NF1 is found at nt 1488 to 1501. Also identified in this search is a region of homology (nt 1708 to 1820) with the internal regulatory element of the quail troponin I gene (40), which is a tissue-specific enhancer located in the first intron of a muscle-specific gene. The region of homology includes the muscle regulatory factor binding site, auxiliary binding sites I and II (41), and a region 3′ to site II. Nucleotides 1744−1751 (Fig. 1) exactly match a direct repeat spanning the muscle factor binding site and site I in the troponin I regulatory element. Overall homology is significant: $p<0.003$.

Expression vectors containing hCMV transcription signals

To secrete heterologous glycoproteins from mammalian cells requires the construction of versatile expression plasmids that take advantage of high-level transient expression in COS cells (28), and that function well when integrated into permanent cell lines. In order to assess the activity of promoter and enhancer functions supplied by hCMV IE1 sequences, the promoterless plasmid pSVori (Fig. 2) was built to contain the SV40 origin of replication, transcript cleavage and polyadenylation signals

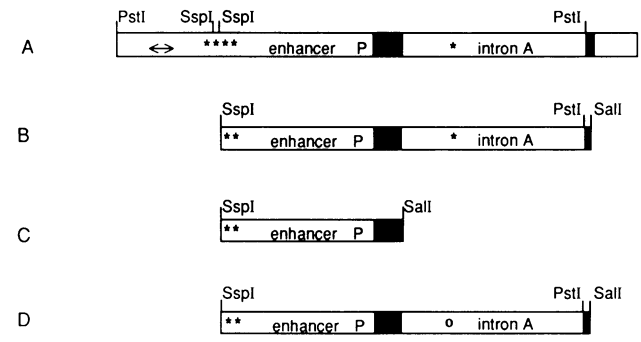

Fig. 3. Fragments of the CMV IE1 regulatory and transcribed regions used in constructions of expression plasmids. *Sal* I sites were introduced synthetically as described in the text. Other restriction sites are found in the sequence shown in Figure 1. Inverted repeats (⟵⟶); NF1 binding sites (*); mutated NF1 binding site (o); promoter (P); exons (■).

from the SV40 T-antigen gene, restriction sites suitable for inserting hCMV DNA fragments, and a prokaryotic selectable marker and origin of replication.

Fig.3 illustrates the hCMV IE1 region used to construct expression vectors for this study. In preliminary experiments, all of the untranscribed regulatory regions of the hCMV IE1 gene, including the potential stem-loop structure spanning nucleotides 119 through 211, clustered NF1 sites, and enhancer/promoter sequences, were included in expression vectors. However, when the first 400 bp of the *Pst* I fragment were present in expression plasmids, poor expression of glycoproteins was observed in both monkey kidney cells (COS7) and in Chinese hamster ovary cells (DXB11). Deletion of these upstream modulatory sequences led

Table I. Expression of HIV-1 gp 160 using enhancer/promoter sequences in CMVIE1 nt 461−2097 compared with the SV40E enhancers/promoter

Plasmid Construction				Expression Level			Reference
Enhancer Promoter	5′ Untrans. Sequence	Signal Peptide	Coding Sequence	Assay Value ng gp120/mL	% of Maximum	No. of Expts.	Plasmid Name
SV40E	SV40/HIV	HIV	gp160	<5	<1.0	2	pSV7cFenv
SV40E	SV40/t-PA	t-PA	gp160	30	6.7	2	pSV7cFenv.tPA
CMVIE1	CMV IE1/t-PA	t-PA	gp160	450	100	2	pCMV6c160

Table II. Effect of intron A on expression of FVIII, HIV and tPA glycoproteins

Plasmid Construction				Expression Level			References
Enhancer Promoter	5′ Untrans. Sequence	Signal Peptide	Coding Sequence	Assay Value	% of Maximum	No. of Expts.	Plasmid Name
				mU FVIII/mL			
SV40E	SV40/FVIII	FVIII	FVIII	13	31	5	pSVF8-302
CMVIE1	IE1Dintron	FVIII	FVIII	31	74	5	pCMVF8-501
CMVIE1	IE1/FVIII	FVIII	FVIII	42	100	10	pCMVF8-600
SV40E	SV40/FVIII	FVIII	FVIIIgp92	106	43	5	pSVF8-92
CMVIE1	IE1Dintron	FVIII	FVIIIgp92	90	37	6	pCMVF8-92/5.1
CMVIE1	IE1/FVIII	FVIII	FVIIIgp92	245	100	8	pCMVF8-92/6x
				ng gp120/mL			
SV40E	SV40/t-PA	t-PA	HIV gp120	<5	<0.3	2	pSV7dARV120tPA
CMVIE1	IE1Dintron/t-PA	t-PA	HIV gp120	40	2.0	2	pCMV5a120SF2
CMVIE1	IE1/t-PA	t-PA	HIV gp120	1930	100	2	pCMV6a120SF2
				ng t-PA/mL			
SV40E	SV40/t-PA	t-PA	t-PA	55.0	100	1	pSV7dtPA3
CMVIE1	IE1Dintron/t-PA	t-PA	t-PA	17.0	31	1	pCMV5atPA3
CMVIE1	IE1/t-PA	t-PA	t-PA	16.1	29	1	pCMV6atPA3

to high levels of expression for several mammalian glycoproteins, suggesting a negative regulatory role for this region in the two cell types. Therefore, glycoprotein expression plasmids were constructed using both transcribed and untranscribed regulatory sequences, including intron A, but lacking 5′ DNA containing the potential stem-loop structure and the first two NF1 binding sites (Fig.3B). Vectors without intron A were prepared using the fragment shown in Fig.3C. To assess the effect of the intronic NF1 binding site on expression, vectors were made using the fragment (Fig.3D) containing a mutation in the NF1 motif in intron A.

These four fragments were inserted into pSVori (Fig.2) to make a set of complete expression plasmids into which cDNAs for human glycoproteins were placed as reporter genes. In comparative experiments described below, we examine expression of human Factor VIII (FVIII), human tissue plasminogen activator (t-PA), and human immunodeficiency virus (HIV-1) coat glycoproteins.

Comparison of the SV40 early and hCMV immediate-early promoter/enhancers

Expression levels from pSVori derivatives containing CMV IE1 promoter/enhancer sequences were compared with expression from the SV40 early promoter/enhancer using HIV-1 envelope glycoproteins as reporter molecules. A cDNA encoding HIV-1 envelope glycoprotein gp160 was subcloned into an expression vector bearing the SV40 early promoter (pSV7d) (24) and into pCMV6, which bears the CMV IE1 enhancer/promoter with intron A (Fig.2). These expression plasmids were transfected into COS7 cells, then supernatants and cell lysates were assayed for the expression of HIV envelope glycoprotein using a sandwich ELISA that recognizes antigens in the gp120 portion of HIV-SF2 gp160. Expression of gp160 is shown in Table I.

HIV-1 gp160 expression was not detected from the SV40 early promoter if the HIV signal sequence was retained (Table I), although expression was detected with human HIV-positive serum using indirect immunofluorescence (data not shown). In an attempt to improve expression, the homologous signal sequence was replaced with that from human t-PA, previously shown to direct high level expression of Factor VIII gp80 (24). [Expression of HIV-1 envelope genes bearing the homologous signal sequence are *rev*-dependent, while genes bearing the t-PA signal are *rev*-independent (N. Haigwood and M.-L. HammarskjÖld, manuscript in preparation). All experiments shown here were done in the absence of *rev* from HIV-1].

Following replacement of the signal sequence with that from the human t-PA gene, expression of gp160 was detectable from the SV40 promoter in COS7 cell lysates (Table I). Expression was still significantly higher from the CMV IE1 enhancer/promoter, at least ten-fold higher than expression driven by the SV40 early enhancer/promoter. Earlier work (11) suggested that expression of t-PA was not significantly higher in a CMV IE1 promoter plasmid than in an SV40 late promoter vector. Results in Table I suggest that the t-PA pre-pro sequence is not sufficient to elevate SV40-driven expression to the level of CMV IE1-driven expression.

Performance of expression vectors with and without intron A

To determine the regulatory role of intron A, transient expression assays were done in the presence and absence of intron A using

Table III. Effect of mutating the NF1 binding site in intron A on expression levels from CMVIE1 plasmids

Plasmid Construction				Expression Level		
NF1 Site Sequence	5′ Untrans. Sequence	Signal Peptide	Coding Sequence	Assay Value mU FVIII/mL	% of Maximum	No. of Expts.
IE1 intron A	IE1/FVIII	FVIII	FVIII gp92	349	100	4
non-binding	IE1/FVIII	FVIII	FVIII gp92	241	69	4
IE1 intron A	IE1/t-PA	t-PA	FVIII gp80	853	100	4
non-binding	IE1/t-PA	t-PA	FVIII gp80	947	112	4
IE1 intron A	co-transf.	both	both	111	100	4
non-binding	co-transf.	both	both	66	59	4

various glycoproteins. Expression plasmids for full-length FVIII, FVIII gp92, HIV-1 gp120 and human t-PA were constructed either in pCMV6 (with intron A) or in pCMV5 (intron A deleted), and compared with SV40 early enhancer/promoter derivatives (Table II).

Four of five glycoproteins were secreted at higher levels when expressed with the CMV IE1 enhancer/promoter than when expressed from the SV40 early enhancer/promoter: FVIII, FVIII gp92, HIV-1 gp160, and HIV-1 gp120 (Tables I and II). With t-PA, the SV40E/t-PA plasmid outperformed both of the CMV IE1/t-PA derivatives. A similar result was obtained in a comparison between the SV40 late promoter and a CMV IE1 construction with t-PA (11) showing that not all glycoprotein cDNAs are best expressed under the control of the CMV IE1 enhancer/promoter. Data in Table II illustrate that expression levels of FVIII, FVIII gp92, and HIV-1 gp120 were significantly improved when intron A was included. Thus, in cases where the CMV IE1 enhancer/promoter is more active than the SV40 enhancer/promoter, intron A contributes to this activity.

The most striking results were obtained with HIV-1 gp120. Using the plasmid without intron A, secretion of gp120 was approximately 2% the level seen from the expression plasmid bearing intron A (Table II). Additional experiments, detecting gp120 by western blotting, have yielded results consistent with these; gp120 is expressed more than 100-fold better using the CMV enhancer/promoter plus intron A compared with the SV40 early promoter (data not shown).

Effect of mutating the NF1 site in intron A

The positive effect of intron A on expression of several glycoproteins may arise by regulation of upstream enhancer/promoter elements (42,43,44,45,46), or by contributions to processing of stable mRNAs (47,48,49). In the transient expression system used to test our plasmid constructions, mRNA stability is probably not a limiting factor. Expression from this system is more likely to be a product of plasmid replication, transcription, and post-translational processing of glycoproteins. Speculating that intron A might contain sequences that function as downstream transcriptional activators (42,43), we searched for several common transcription elements. The sequence of intron A contains only one of these elements, a consensus binding site for nuclear factor 1 (14,15).

Experiments shown in Table III were done to determine whether binding of NF1 to the site in intron A was responsible for some or all of its positive effects. An expression plasmid was constructed using a fragment (Fig.3D) in which the consensus NF1 binding site in intron A was mutated to a sequence reported not to bind NF1 (14). Derivative plasmids for expression of FVIII gp92 and gp80 were cloned, transfected into COS7 cells, and assayed for FVIII activity. Data in Table III show that changing the NF1 motif in intron A had no effect on expression of FVIII gp80 (a construction having 5′ and 3′ sequences derived from t-PA), but consistently reduced expression of FVIII gp92; co-transfection of the gp80 and gp92 expression plasmids confirmed the results with gp92 alone. The decrease in FVIII gp 92 expression of 30−40% was not as great as the 60% observed after deletion of intron A (Table III).

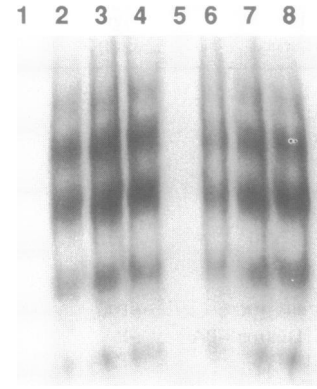

Fig. 4. Detection of NF1 in COS7 nuclear extracts. The assay was performed as described in Materials and Methods in 20 μL mixtures containing binding buffer (lanes 1 and 5), 1 μL of COS7 nuclear extract (lanes 2 and 6), 2 μL nuclear extract (lanes 3 and 7), or 4 μL nuclear extract (lanes 4 and 8). Mixtures included 10 fmol ^{32}P-labeled intron A NF1 oligonucleotide (lanes 1−4) or mutant oligonucleotide (lanes 5−8). Five major complexes are detected in lanes where nuclear extracts are present. Unbound oligonucleotide was run off the gel.

NF1 binding to its consensus motif TGGC/AN$_5$GCCAA has been shown to be sensitive to the sequence of flanking DNA (37). While the NF1 sequence represented in our mutant NF1 site does not function as an NF1 binding site in the middle repetitive element of mouse DNA (14), NF1 could bind this sequence in the context of intron A. Moreover, the composition of the five bp spacer can modulate binding affinity and expression activity over a four-fold range (39). Data shown in Table III would be consistent with the hypothesis that the mutant NF1 site in our test plasmid binds NF1, but the three mutations in the spacer reduce NF1/DNA binding affinity, leading to reduced expression.

To test this hypothesis, we performed a gel mobility retardation assay using crude nuclear extracts from COS7 cells as a source of NF1 proteins. These proteins were allowed to bind labeled, double-stranded oligonucleotides representing the context and sequence of the intron A and mutant NF1 sites. This assay shows

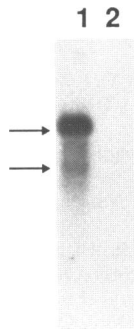

Fig. 5. Autoradiogram of total cytoplasmic RNA from a pCMV6 plasmid. RNA was isolated from permanent ts COS cell lines grown at 33°, electrophoresed and blotted onto nitrocellulose, and probed by hybridization to a ^{32}P-labeled restriction fragment of HIV-1 gp 120. Lane 1, tsCOS-129 (co-transfected with the gp120 expression plasmid and neomycin resistance plasmid); lane 2, tsCOS-N6 (transfected with neomycin resistance plasmid).

differential migration of protein-bound oligonucleotides in a non-denaturing polyacrylamide gel (Fig.4). COS7 nuclear extracts clearly contain several NF1 proteins capable of binding the mutant NF1 site *in vitro* (Fig.4, lanes 6−8). However, equivalent amounts of nuclear extract bind fewer molecules of the mutant oligomer than intron A oligomer (Fig.4, lanes 2 vs 6, etc.). This reduction in NF1 binding to the mutant sequence was observed to be between two and five fold in repeated assays (data not shown). The difference seen in NF1 binding to the mutant oligonucleotide *in vitro* is consistent with the lower level of expression of FVIII gp92 from the mutant expression vector. We conclude from these experiments that binding of NF1 to the site in intron A can increase expression directed by the CMV IE1 enhancer/promoter, but probably does not account for all of the positive effects of intron A.

Steady state mRNA transcribed from intron A-containing plasmids

To examine the quantity and size of transcripts produced from pCMV6-derived plasmids, steady state mRNA was analyzed in several HIV-1 gp 120-expressing permanent cell lines. Temperature sensitive (ts) COS cells (31) were grown under permissive (33°C) conditions to induce transcription, and total cytoplasmic RNA was examined on Northern blots hybridized with a HIV-1 gp120-specific probe. Shown for one such cell line (Fig. 5), the predominant species is larger than predicted for a spliced transcript polyadenylated at the SV40 early site (upper arrow, ~2.6 kb). A minor species appears at the size expected for correctly spliced and polyadenylated mRNA (lower arrow, ~1.6 kb). The same sizes and proportions of mRNA were observed in these cell lines at the non-permissive (40°C) temperature, but at lower levels (data not shown).

The larger (predominant) cytoplasmic species probably results from read-through of the weak SV40 polyadenylation signal supplied by the vector, with cleavage and polyadenylation occurring at a cryptic site approximately one Kb downstream in the prokaryotic β-lactamase gene (50,51). Cytoplasmic RNA transcribed from construct pSVF8-302 (see Table II), which has the same polyadenylation signal and downstream prokaryotic sequences, hybridizes with pBR322 DNA (data not shown).

It has been suggested that one effect of introns in mammalian expression vectors may be to improve the efficiency of RNA processing and transport to the cytoplasm (49). The large amount of cytoplasmic message found in cells transfected with intron A-containing plasmids is consistent with this function.

DISCUSSION

While untranscribed regions of highly-expressed genes have been extensively studied, the role of transcribed sequences has only recently been appreciated. Our study suggests that intron A of the hCMV IE1 gene activates its promoter in transformed monkey kidney cells. In earlier work, including introns in expression plasmids was found to increase levels of expression in both transient systems and permanent cell lines (47,49,52). Transcriptional activation (42,43,44,45,46) and mRNA stabilization (47,48) have been suggested as possible mechanisms underlying this effect.

Recent work indicates that regulatory elements in transcribed sequences may be integral regulators of promoters (17,42,43,53). Strong enhancer-like elements have been found in transcribed sequences of many genes, *e.g.*, the enhancer found in the intron between the J_H and μ constant region exons of immunoglobulin heavy chain genes (53,54). More regulatory elements occur in the first introns of structural genes such as the mouse ribosomal protein gene rpL32 (45), the mouse alpha 1 type I collagen gene (46), and the chicken troponin I gene (40,41). These elements appear to be promoter-specific.

In the present case, intron A of the hCMV IE1 gene seems to contain only one consensus site for a well-characterized transcription factor, an NF1 binding site. Our data show that NF1 binding to this site has a small effect on the activity of the hCMV IE1 enhancer/promoter. A search of intron A for enhancer-like elements reveals homology with the internal regulatory element of the troponin I gene. However, since our assays were performed in non-muscle cells, this enhancer-like element could be functioning by binding related transcription factors present in transformed kidney cells. When a pCMV6 (intronA-containing) Factor VIII plasmid was introduced into transgenic mice, FVIII glycoprotein was expressed primarily in muscle tissues (T.Mikkelsen, B.Chapman, N.Din, J.Ingerslev, P.Kristensen, K.Poulsen, and J.P.Hjorth, in preparation). This tissue specificity was not seen in transgenic mice when intron A was absent from the CMV IE1 sequences (10). These results indicate that sequences in intron A, which are homologous to muscle-specific regulatory elements, may be functioning in a tissue-specific manner *in vivo*.

Not all cDNAs behave the same way in these vectors. We conclude that no host/vector combination is likely to be universal for good glycoprotein expression, and emphasize that the results presented above do not provide data on relative levels of transcription or mRNA stability. Instead, we have assessed the role of the CMV IE1 regulatory region in a complex, multi-step process. Expression of glycoproteins, unlike artificial reporter molecules, may be limited less by transcription and more by protein folding, glycosylation, and other post-translational processes. Positive effects of intron A have been demonstrated in a transient, high copy number system, and recently in permanent hamster cell lines. The CMV IE1 enhancer/promoter with intron A has produced very high levels of glycoproteins in a system where mRNA stability contributes to robust expression (Chapman and Haigwood, unpublished). Our results indicate that

heterologous expression vectors can be optimized by inclusion of intron A with other elements of the hCMV IE1 enhancer/promoter.

ACKNOWLEDGEMENTS

The authors would like to thank Nicholas Chester, Patricia Scott, Julia Carlson, Gregory Moore, Bruce Irvine, Kimberly Mann, Christine Evans, Sara XingZi Wu, and Lynda Misher for excellent technical assistance. We would also like to thank Dr. Paul Luciw for generously providing plasmids pSV7d, pAd-DHFR, and pSV1-DHFR. We are grateful to Dr. Rae Lyn Burke for suggesting the use of an *Nhe* I site in the t-PA signal sequence, and to Dr. Alan M. Weiner for helpful comments on the manuscript.

REFERENCES

1. Boshart, M., Weber, F., Jahn, G., Dorsch-Hasler, K., Fleckenstein, B. and Schaffner, W. (1985) *Cell*, **41**, 521–530.
2. Thomsen, D.R., Stenberg, R.M., Goins, W.F. and Stinski, M.F. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 659–663.
3. Ghazal, P., Lubon, H., Fleckenstein, B. and Hennighausen, L. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 3658–3662.
4. Foecking, M.K. and Hofstetter, H. (1986) *Gene*, **45**, 101–105.
5. Sevarino, K.A., Felix, R., Banks, C.M., Low, M.J., Montminy, M.R., Mandel, G. and Goodman, R.H. (1987) *J. Biol. Chem.*, **262**, 4987–4993.
6. Pasleau, F., Tocci, M.J., Leung, F. and Kopchick, J.J. (1985) *Gene*, **38**, 227–232.
7. Cockett, M.I., Bebbington, C.R. and Yarranton, G.T. (1990) *Biotechnology*, **8**, 662–667.
8. Andersson, S., Davis, D.L., Dahlback, H., Jornvall, H. and Russell, D.W. (1989) *J. Biol. Chem.*, **264**, 8222–8229.
9. Eaton, D.L., Wood, W.I., Eaton, D., Hass, P.E., Hollingshead, P., Wion, K., Mather, J., Lawn, R.M., Vehar, G. and Gorman, C. (1986) *Biochemistry*, **25**, 8343–8353.
10. Schmidt, E.V., Christoph, G., Zeller, R. and Leder, P. (1990) *Mol. Cell. Biol.*, **10**, 4406–4411.
11. Stephens, P.E. and Cockett, M.I. (1989) *Nucleic Acids Res.*, **17**, 7110.
12. Lubon, H., Ghazal, P., Hennighausen, L., Reynolds-Kohler, C., Lockshin, C. and Nelson, J. (1989) *Mol. Cell. Biol.*, **9**, 1342–1345.
13. Nelson, J.A., Reynolds-Kohler, C. and Smith, B.A. (1987) *Mol. Cell. Biol.*, **7**, 4125–4129.
14. Hennighausen, L. and Fleckenstein, B. (1986) *EMBO J.*, **5**, 1367–1371.
15. Goyal, N., Knox, J. and Gronostajski, R.M. (1990) *Mol. Cell. Biol.*, **10**, 1041–1048.
16. Stenberg, R.M., Thomsen, D.R. and Stinski, M.F. (1984) *J. Virol.*, **49**, 190–199.
17. Ghazal, P. and Nelson, J.A. (1991) *J. Virol.*, **65**, 2299–2307.
18. LaFemina, R.L. and Hayward, G.S. (1980) In Fields, B.N. and Jaenish, R. (ed.), Animal Virus Genetics. Academic Press, New York, Vol.**18**, pp. 39–55.
19. Hanahan, D. (1983) *J. Mol. Biol.*, **166**, 557–567.
20. Messing, J. (1983) In Wu, R., Grossman, L. and Moldave, K. (ed.), Methods in Enzymology. Academic Press, New York, Vol.**101**, pp. 20–78.
21. Zoller, M.J. and Smith, M. (1983) In Wu, R., Grossman, L. and Moldave, K. (ed.), Methods in Enzymology. Academic Press, New York, Vol.**100A**, pp. 468–500.
22. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467.
23. Rio, D.C. and Tjian, R. (1983) *Cell*, **32**, 1227–1240.
24. Burke, R.L., Pachl, C., Quiroga, M., Rosenberg, S., Haigwood, N., Nordfand, O. and Ezban, M. (1986) *J. Biol. Chem.*, **261**, 12574–12578.
25. Haigwood, N.L., Mullenbach, G.T., Moore, G.K., Desjardin, L.E., Trabrizi, A., Brown-Shimer, S.L., Stauss, H., Stöhr, H.A. and Pâques, E.-P. (1989) *Prot. Eng.*, **2**, 611–620.
26. Barr, P.J., Thayer, R.M., Laybourn, P., Najarian, R.C., Seela, F. and Tolan, D.R. (1986) *Biotechniques*, **4**, 428–432.
27. Martinez, H. (1985) Analysis Release Package. Biomathematics Computation Laboratory, Department of Biochemistry and Biophysics, University of California, San Francisco.
28. Gluzman, Y. (1981) *Cell*, **23**, 175–182.
29. Sompayrac, L.M. and Danna, K.J. (1981) *Proc. Natl. Acad. Sci. USA*, **78**, 7575–7578.
30. Luthman, H. and Magnusson, G. (1983) *Nucleic Acids Res.*, **11**, 1295–1308.
31. Rio, D.C., Clark, S.G. and Tjian, R. (1985) *Science*, **227**, 23–28.
32. Southern, P.J. and Berg, P. (1982) *J. Mol. Appl. Genet.*, **1**, 327–341.
33. Nordfang, O., Ezban, M. and Dinesen, B. (1985) *Thromb. Haemost.*, **53**, 346–350.
34. Nordfang, O., Ezban, M., Nilsson, P. and Knudsen, J.B. (1988) *Br. J. Haematol.*, **68**, 307–312.
35. Haigwood, N.L., Barker, C.B., Higgins, K.W., Skiles, P.V., Moore, G.K., Mann, K.A., Lee, D.R., Eichberg, J.W. and Steimer, K.S. (1990) In Brown, F., Chanock, R.M., Ginsberg, H.S. and Lerner, R.A. (ed.), Vaccines 90. Cold Spring Harbor Laboratory, Cold Spring Harbor, pp. 313–320.
36. Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) Molecular Cloning. Cold Spring Harbor Laboratory Press, Cold Spring Harbor.
37. Gronostajski, R.M. (1986) *Nucleic Acids Res.*, **14**, 9117–9132.
38. Akrigg, A., Wilkinson, G.W.G. and Oram, J.D. (1985) *Virus Res.*, **2**, 107–121.
39. Knox, J.J., Rebstein, P.J., Manoukian, A. and Gronostajski, R.M. (1991) *Mol. Cell. Biol.*, **11**, 2946–2951.
40. Yutzey, K.E., Kline, R.L. and Konieczny, S.F. (1989) *Mol. Cell. Biol.*, **9**, 1397–1405.
41. Lin, H., Yutzey, K.E. and Konieczny, S.F. (1991) *Mol. Cell. Biol.*, **11**, 267–280.
42. Ayer, C.E. and Dynan, W.S. (1990) *Mol. Cell. Biol.*, **10**, 3635–3645.
43. Mansour, S.L., Grodzicker, T. and Tjian, R. (1986) *Mol. Cell. Biol.*, **6**, 2684–2694.
44. Ghogawala, Z., Choi, E., Daly, K.R., Blanco, L.R., Griffith, I.J. and Glimcher, L.H. (1989) *Mol. Cell. Biol.*, **9**, 4402–4408.
45. Chung, S. and Perry, R.P. (1989) *Mol. Cell. Biol.*, **9**, 2075–2082.
46. Rippe, R.A., Sven-Iver, L., Brenner, D.A. and Breindl, M. (1989) *Mol. Cell. Biol.*, **9**, 2224–2227.
47. Buchman, A.R. and Berg, P. (1988) *Mol. Cell. Biol.*, **8**, 4395–4405.
48. Hamer, D.H. and Leder, P. (1979) *Cell*, **18**, 1299–1302.
49. Huang, M.T.F. and Gorman, C.M. (1990) *Nucleic Acids Res.*, **18**, 937–947.
50. Zhang, F., Denome, R.M. and Cole, C.N. (1986) *Mol. Cell. Biol.*, **6**, 4611–4623.
51. Denome, R.M. and Cole, C.N. (1988) *Mol. Cell. Biol.*, **8**, 4829–4839.
52. Gorman, C.M., Rigby, P.W.J. and Lane, D.P. (1985) *Cell*, **42**, 519–526.
53. Imler, J.-L., Lemaire, C., Wasylyk, C. and Wasylyk, B. (1987) *Mol. Cell. Biol.*, **7**, 2558–2567.
54. Banerji, J., Olson, L. and Schaffner, W. (1983) *Cell*, **33**, 729–740.